

A new generation of nanobody research tools using improved mass spectrometry-based discovery methods

Received for publication, March 13, 2024, and in revised form, July 1, 2024. Published, Papers in Press, August 2, 2024.
<https://doi.org/10.1016/j.jbc.2024.107623>

Peter C. Fridy¹, Ryan J. Farrell^{2,3}, Kelly R. Molloy⁴, Sarah Keegan⁵, Junjie Wang⁴, Erica Y. Jacobs^{4,6}, Yinyin Li⁴, Jill Trivedi¹, Viren Sehgal¹, David Fenyo⁵, Zhuhao Wu², Brian T. Chait^{4,*}, and Michael P. Rout^{1,*}

From the ¹Laboratory of Cellular and Structural Biology, ²Laboratory of Brain Development and Repair, The Rockefeller University, New York, New York, USA; ³Department of Biochemistry, Weill Cornell Medicine, New York, New York, USA; ⁴Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York, USA; ⁵Department of Biochemistry and Molecular Pharmacology, Institute for Systems Genetics, NYU Grossman School of Medicine, New York, New York, USA; ⁶Chemistry Department, St John's University, Queens, New York, USA

Reviewed by members of the JBC Editorial Board. Edited by Philip A. Cole

Single-domain antibodies (“nanobodies”) derived from the variable region of camelid heavy-chain only antibody variants have proven to be widely useful tools for research, therapeutic, and diagnostic applications. In addition to traditional display techniques, methods to generate nanobodies using direct detection by mass spectrometry and DNA sequencing have been highly effective. However, certain technical challenges have limited widespread application. We have optimized a new pipeline for this approach that greatly improves screening sensitivity, depth of antibody coverage, antigen compatibility, and overall hit rate and affinity. We have applied this improved methodology to generate significantly higher affinity nanobody repertoires against widely used targets in biological research—*i.e.*, GFP, tdTomato, GST, and mouse, rabbit, and goat immunoglobulin G. We have characterized these reagents in affinity isolations and tissue immunofluorescence microscopy, identifying those that are optimal for these particularly demanding applications, and engineering dimeric constructs for ultra-high affinity. This study thus provides new nanobody tools directly applicable to a wide variety of research problems, and improved techniques enabling future nanobody development against diverse targets.

Nanobodies are single-domain, antigen-specific proteins derived from the variable domain (V_HH) of a heavy chain only immunoglobulin G (IgG) antibody variant (HCAb) found in camelid species such as llamas, alpacas, and camels (1). Though approximately one-tenth the size of typical antibodies, nanobodies retain comparable antigen affinity and specificity, can access novel antigenic sites, and can be readily cloned and expressed recombinantly in bacteria (2–4). In comparison with monoclonal antibodies, they possess numerous important

advantages, including increased stability, tissue penetrance, solubility, and high amenability to genetic or chemical engineering (5, 6). These advantages have made nanobodies an increasingly popular option for therapeutic, diagnostic, and research applications that demand targeted high-affinity reagents (7, 8).

Nanobodies can typically be generated from immunized camelids through phage display technology (9, 10), yeast display (11–13), or through direct identification by mass spectrometry (MS) combined with high-throughput DNA sequencing (14–16). The MS-based pipeline for nanobody identification has been successfully used to generate large repertoires of high-quality reagents, and has significant advantages in stringency, speed, and lack of an intermediate library; however, certain technical demands have limited the applications and widespread adoption of this approach. We therefore sought to significantly improve the ease of this technology as well as the quality of its resulting nanobody candidates, allowing it to be applied with high success rates to a much wider range of target antigens. We made improvements throughout our pipeline including in the methods used for serum antibody sample preparation, in the MS analysis of the resulting purified V_HH mixtures, and in the bioinformatics software used for candidate identification and ranking. We also significantly improved the coverage of V_HH sequences from lymphocyte complementary DNA (cDNA) by optimizing the primers used for PCR amplification; this optimization is widely applicable to nanobody identification approaches, as V_HH amplification is also a key step in preparing display libraries from either naive or immunized animals.

As examples from our improved MS-based method, we report new or improved nanobodies against several antigens of broad research use, including the fluorescent protein tags GFP and tdTomato, the expression fusion protein GST tag, as well as IgG from mouse, rabbit, and goat. These nanobodies are ultra-high affinity, with K_Ds into the sub-nanomolar range, and are suitable for high-quality immunoprecipitation and immunofluorescence (IF) applications.

* For correspondence: Michael P. Rout, rout@rockefeller.edu; Brian T. Chait, chait@rockefeller.edu.

Present address for: Ryan J. Farrell, Neuroscience Institute, NYU Medical Center, New York, NY, USA; Zhuhao Wu, Appel Alzheimer's Disease Research Institute, Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY, USA.

Nanobody research tools using improved MS-based discovery

Results

Our new methods for nanobody identification build off our previously published pipeline and rely on similar parallel procedures (Fig. 1). First, the variant HCAB of interest is purified from crude serum taken from an immunized camelid. The target antigen is then immobilized and used to affinity-isolate antigen-specific HCAB from this fraction. Bound HCAB is protease-treated on the antigen resin to release the Fc domains, and the remaining bound $V_{H}H$ is eluted and resolved by SDS-PAGE. $V_{H}H$ -specific proteins are digested and analyzed by liquid chromatography tandem MS to generate peptide sequence data. In parallel to this serum analysis, RNA is extracted from lymphocytes purified from the bone marrow sampled from the same animal. Expressed $V_{H}H$ sequences from this RNA are then amplified by RT-PCR to produce a cDNA library for Illumina sequencing. Antigen-specific nanobodies are identified by using our Llama-Magic software

(https://github.com/FenyoLab/Llama_Nanobodies) to search the library with the peptide sequence data. All steps of these parallel procedures were optimized from their original versions for improved nanobody identification and robustness to sample variability.

Improvements in the identification of nanobodies from llama sera

Direct identification of nanobody candidates from camelid serum relies on MS analysis of highly purified antigen-specific HCAB protein. Previously, HCAB fractions have been obtained from crude serum by selective binding to and elution from Protein A and Protein G resin. While this results in significant enrichment of HCAB, IgG contamination is typical in this preparation (5–15%) (14), interfering with MS analysis by introducing irrelevant peptides and complexity. We took advantage of an additional IgG-binding factor from

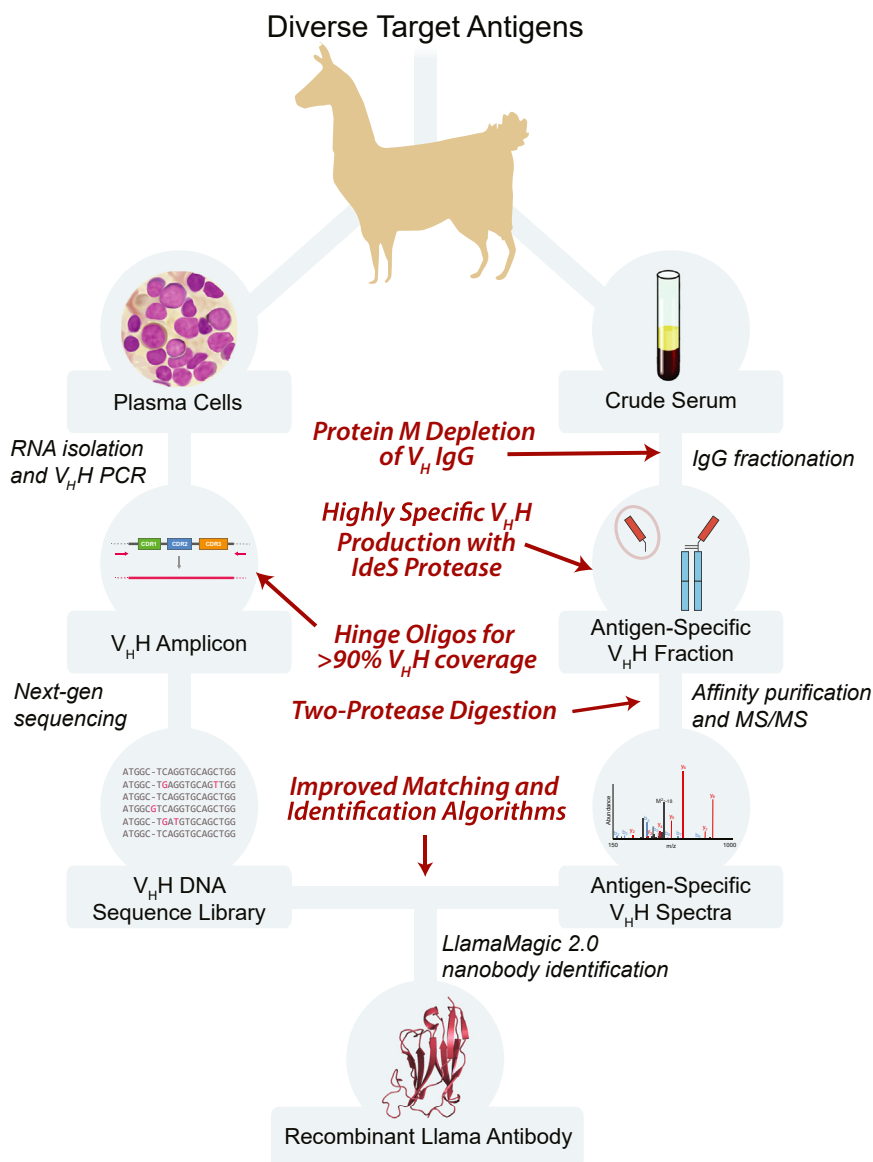


Figure 1. Improved pipeline for nanobody generation by MS and next-gen DNA sequencing. Overview of the methodology used to identify nanobody candidates from immunized llamas, with major improvements indicated in red. MS, mass spectrometry.

mycoplasma, protein M, which specifically targets light chain subunits and should therefore bind standard IgG but not HCAb (17). As predicted, passing our samples over protein M resin led to virtually complete depletion of IgG (Fig. 2A). By incorporating this depletion step into sample preparation upstream of MS, we essentially eliminate this source of contamination.

Another key concern in the quality of protein purified for MS is the removal of Fc, which otherwise increases spectral complexity and complicates V_{HH} peptide identification. This was initially done by a traditional papain digest of HCAb captured on antigen resin. While this effectively removes Fc, papain is a relatively nonspecific protease, and depending on the antigen used for purification it can be difficult or impossible to tune digestion conditions to cleave off Fc without excessive degradation of either the immobilized antigen or the V_{HH} fragment itself. As an alternative, we surveyed sequence-specific proteases and identified the streptococcal protease IdeS as a superior option for Fc/F(ab')₂ digestion of llama or alpaca HCAb (18). Compared to papain across a variety of protein standards, IdeS digestion readily cleaves IgG with no off-target proteolysis (Fig. S1A). Moreover, when used on affinity-purified llama HCAb, IdeS treatment produces distinct bands correlating to Fc and V_{HH} domains which are readily separated by SDS-PAGE for MS analysis (Fig. 2A). In comparing MS samples prepared from serum previously used to identify anti-GFP nanobody candidates (14), both IdeS and papain performed equally better than no Fc digestion, but IdeS without concomitant off-target digestion issues (Fig. S1, A and B).

Optimization of V_{HH} mass spectrometric analysis

Nanobody sequences are identified by computationally matching peptides to full-length protein sequences identified from *in silico* translation of cDNA, and so adequate peptide coverage is vital to high-confidence candidate selection. Peptides covering the complementarity determining region 3 (CDR3) are particularly important, as this is the most divergent and defining region of each species of V_{HH} (19). Multiple improvements were thus made to our MS analyses to improve peptide identification and sequence matching, as follows. First, SDS-PAGE-resolved and Coomassie blue-stained V_{HH} bands were excised from the separating gel, split in half, and each digested in parallel with either trypsin or chymotrypsin, rather than with trypsin alone. As previously observed, regions adjacent to CDR3 were more favored for cleavage by chymotrypsin than trypsin (Fig. 2B) (15). Trypsin nevertheless provided superior V_{HH} coverage overall, and we found that integrating MS data from parallel trypsin and chymotrypsin samples provided the most comprehensive sequence coverage (Fig. S1C).

Additionally, we introduced an optional subslicing of SDS-PAGE separated V_{HH} bands before MS to enable greater correlation between peptides derived from a single V_{HH} protein. A major challenge of MS identification of sequences from a complex antibody mixture is the need to identify and correlate the three associated CDRs, separated by highly conserved framework regions, to define a unique antibody

sequence; thus, for high confidence identification, peptides covering the different CDRs must be correctly matched to a single origin protein. To improve this correlation, the V_{HH} gel region can be sliced horizontally to generate several sections for digestion and MS analysis, significantly decreasing the complexity of each sample and limiting incorrect peptide matches across different V_{HH} proteins (Fig. S1D).

In cases where affinity isolations have yielded several micrograms of V_{HH} protein, we have also used offline high pH reversed phase peptide fractionation to decrease the complexity of each sample subjected to mass spectrometry. With a decrease in the chromatographic density of each fraction and an increase in the overall amount of sample that can be analyzed by LC-MS, fractionation enhances the sensitivity of peptide identification and therefore improved V_{HH} sequence coverage, as previously established in other MS workflows (20).

Using RNA-seq to redesign PCR amplification of V_{HH} from lymphocyte cDNA

Previous V_{HH} cDNA libraries used for high-throughput sequencing or display approaches to nanobody selection have generally relied on primer sets complementary to the relatively conserved FR1 and FR4 framework regions of the V_{HH} domain (10, 21, 22). However, in other studies, we have found surprising amounts of divergence in these framework regions (4, 14), and so sought to evaluate the V_{HH} diversity that our existing primer sets were able to amplify.

Given the higher conservation outside of the V_{HH} domain, we first performed long read PacBio sequencing of PCR products amplified from llama or alpaca lymphocyte cDNA by primers annealing to the flanking IgG leader (termed primer CALL001) and CH2 (CALL002) regions, as typically used in the first step of nested PCRs amplifying V_{HH} (Fig. S2A) (21). V_{HH} DNA was separated from longer V_H fragments (containing CH1) by agarose gel electrophoresis. We analyzed the first 18 bp of V_{HH} domains from 6353 llama sequences and found that the top 20 most prevalent 5' sequences came from only 47% of the population and contained up to five substitutions from the consensus sequence (Fig. S2B). In alpaca, 41% were covered by the top 20. This suggested that many divergent primers would be necessary to achieve representative amplification of the V_{HH} population. Similar analysis of the 3' end revealed higher conservation, with the top 20 sequences covering 92% of the population in both species, with three or fewer substitutions in this set.

To avoid potential PCR bias and assess sequences not amplified by CALL001/CALL002, we also surveyed the V_{HH} sequence population by direct Illumina HiSeq sequencing of llama lymphocyte cDNA, resulting in 33 million paired-end reads. We first sought to determine whether the CALL001 primer annealing in the IgG leader sequence achieved adequate V_{HH} coverage. To identify target IgG regions from sequencing reads, we searched for matches to the CALL001 sequence, allowing 20% mismatches (5 in 23 bp) to ensure less conserved sites were captured, without excessive off-target matching. The top ten most abundant sequences were

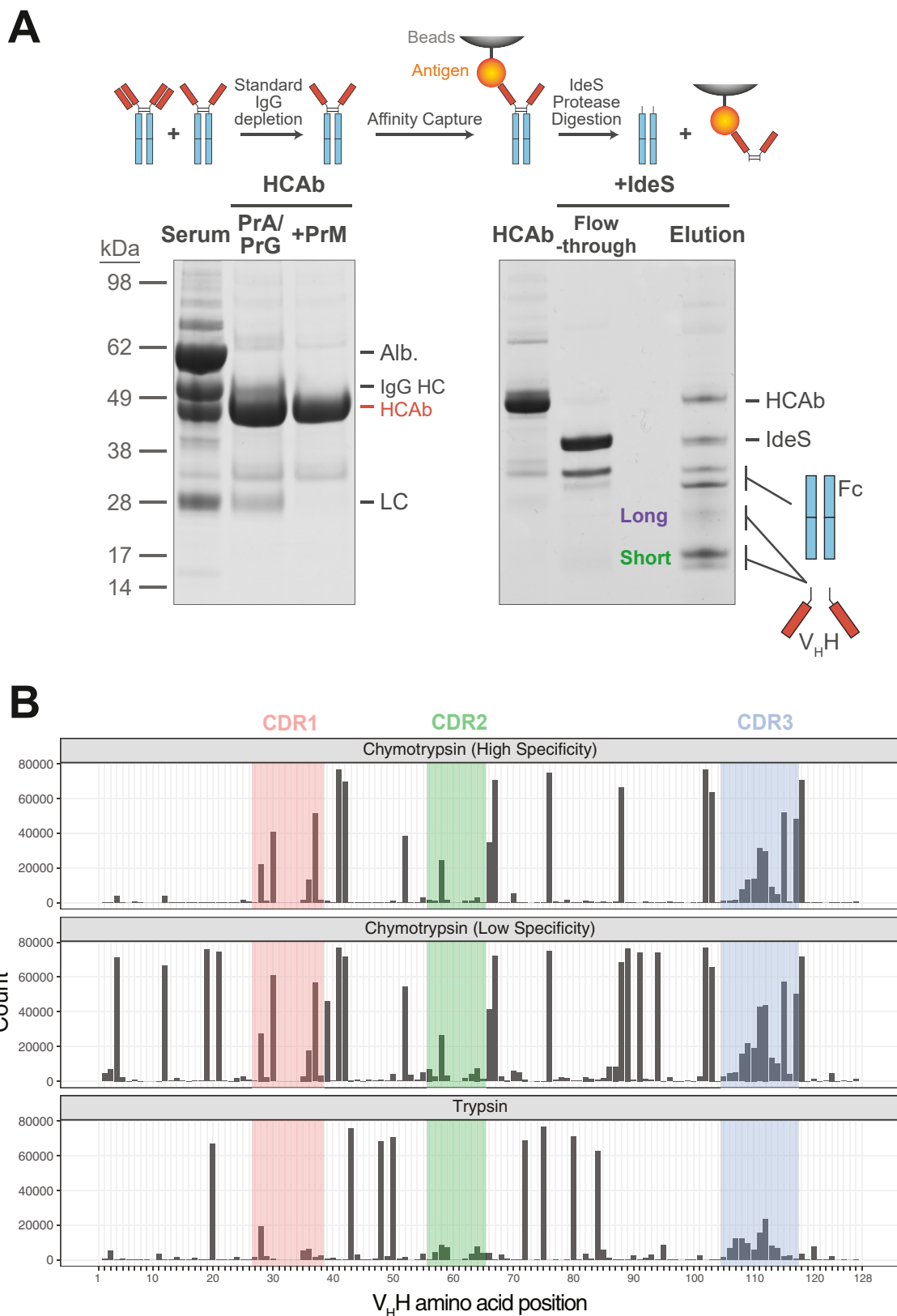


Figure 2. Improvements in V_HH isolation and digestion for MS. A, representative purification of HcAb using Protein A, G, and M (left), and IdeS digestion after affinity isolation (right). Coomassie-stained SDS-PAGE gels are shown, with bands labeled based on dominant protein as identified by MS. Crude serum input ("Serum") was purified over Protein A and Protein G for initial HcAb enrichment ("PrA/PrG") at 94 ± 0.8% purity (n = 5). A second step depletion over PrM ("PrM") results in a high purity HcAb fraction: HcAb yield was 89 ± 2.2%, with standard IgG not measurable above background (n = 5). After affinity capture of HcAb on antigen resin (right), IdeS was added to digest bound HcAb. After digestion, cleaved Fc and IdeS are released in the reaction solution ("Flow-through"). Remaining antigen-bound fragments are then eluted by boiling in SDS ("Elution"). B, *in silico* protease digestion was performed against an experimental llama V_HH sequence library. Counts of protease cleavage events at a given residue are plotted, using IMGT IgG amino acid numbering, with

selected, with less abundant sequences merged by allowing one mismatch to account for potential sequencing errors or point mutations (Fig. S2C). This top ten covered 87% of the identified sequences, including all those with IgG flanking regions consistent with genomic sequences (23). This analysis suggests that CALL001 indeed targets a highly conserved sequence in the IgG leader region suitable for comprehensive V_HH amplification, with significantly higher conservation than alternatives targeting FR1.

While existing CH2 and FR4 3' primers did appear to target well-conserved sequences, these present limitations for amplifying the V_HH domain: first, use of FR4 primers necessitates a less efficient nested PCR strategy to separate V_H and V_HH, as FR4 is conserved between these variants; and second, for purposes such as MS, it is useful to have sequence information into the flanking hinge region, as this allows better identification of CDR3 peptides extending into the hinge, and identifies the corresponding hinge variant (IgG2 or IgG3) for each sequence. While use of the downstream CALL002 CH2 primer would address these concerns, it would be too distant from the V_HH region for typical library or sequencing approaches. We therefore aimed to design primers targeting the flanking V_HH hinge regions.

We identified candidate conserved sites in both long and short (IgG2 and IgG3) hinge regions immediately downstream of the V_HH domain (Fig. 3A). Searching these candidate primer sites against cDNA sequencing reads and allowing up to 20% mismatches, we again merged matches into the top ten most abundant sequences. These accounted for 94% or 91% in long or short hinge samples, respectively (Figs. 3B and S2C). With less than one mismatch within each top ten set, these primer candidates are suitable for comprehensive amplification of V_HH domains. Analysis of adjacent FR4 sequences also revealed significantly more diversity in this framework region when amplified using hinge primers, again indicative of the greater coverage achieved when targeting outside of the V_HH domain (Fig. 3C).

The combination of leader and hinge primers allows comprehensive, single-step PCR amplification of V_HH from cDNA. While this adds a further 27 bp upstream and 22/23 bp downstream of the V_HH domain compared to FR1/FR4 primers, this did not interfere with sequencing or expression. These primers are also compatible with display library approaches from naïve or immunized camelids, and should similarly allow for increased V_HH coverage of libraries generated for these methods by PCR amplification.

Improved workflow for nanobody sequence identification

The Llama-Magic software that we previously used to correlate MS data to sequenced V_HH transcripts was substantially updated to improve the confidence of nanobody candidate selection (14). This improved version is now available as Llama-Magic 2.0 (https://github.com/FenyoLab/Llama_Nanobodies). Merging of paired end reads was

modified to account for new PCR amplicon composition, and *in silico* digestion can be performed with both trypsin and chymotrypsin, with peptides matched to the corresponding digested sample. Sequences are scored by the degree of peptide coverage with additional weight being given to CDR coverage. The resulting candidates are grouped by CDR3 sequence, where CDR3 groups are ranked by their top-scoring members. Then, sequences within CDR3 groups are ranked by high-throughput sequencing counts; given the inherent errors in DNA sequencing and PCR fidelity, higher count sequences are more likely to be from *bona fide* transcripts, and have indeed correlated more strongly with true antigen binders. Additionally, we have found that increasing Illumina sequencing coverage above ~10 million reads does not significantly increase the quality of top nanobody matches, suggesting that at least in this pipeline, rarer clones are not an essential component of the captured high-affinity serum HCAs.

Each peptide identified by mass spectrometry is assigned a “uniqueness score” for each CDR3 group to which it matches, calculated as 100 times the number of sequences in the CDR3 group that contain the peptide, divided by the number of sequences in the entire database that contain the peptide. A peptide’s uniqueness score is therefore the probability that it originated from an antibody with the given group’s CDR3. Selection of a candidate sequence requires at least one peptide with a high uniqueness score (>80). Typically, peptides containing the CDR3 are the most unique within a given V_HH sequence. The MS/MS spectra of these peptides are examined to confirm fragmentation within any part of the CDR3 they may span. The same logic is applied to peptides that span parts of CDR1 and 2. The likelihood that a sequence will ultimately bind to the target antigen is increased by unique peptide coverage across its length.

Candidate lists generated from Llama-Magic 2.0 output were manually filtered with the following additional criteria. First, sequences that were identified in V_HH samples purified against unrelated antigens were eliminated as likely arising from nonspecific binding, and sequences containing the GLEW motif at residues 49 to 52 (IMGT numbering) were also removed as likely VH contaminants (24). Candidates passing these filters were selected for gene synthesis, expression as nanobodies, and antigen-binding screening (14). With the increased stringency of this candidate identification and upstream sample preparation, the false discovery rate is substantially lower than previous results: the overall percentage of selected candidates with verified antigen binding increased from 57% (37/65) using our prior method (14) to 76% (69/91) in this report, and is routinely as high as 90% with other antigens (4).

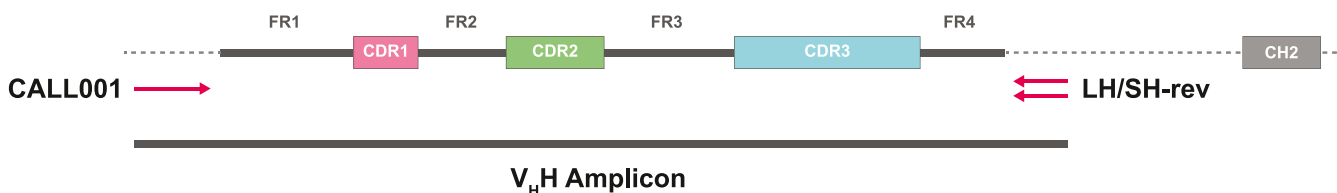
Nanobody repertoires obtained with our improved MS-based methodology

To validate the improvements made to our nanobody generation pipeline, we applied it to several target antigens of

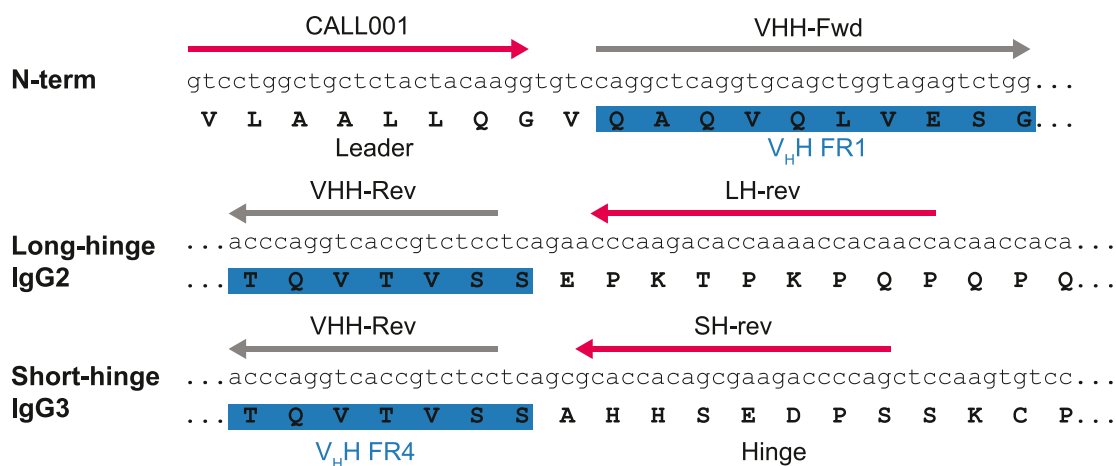
CDRs highlighted in *color*. For chymotrypsin, “low-specificity” predictions include potential cleavages after Met, Leu, and His residues, in addition to “high-specificity” cleavages after Phe, Tyr, and Trp. CDR, complementarity determining region; IgG, immunoglobulin G.

A

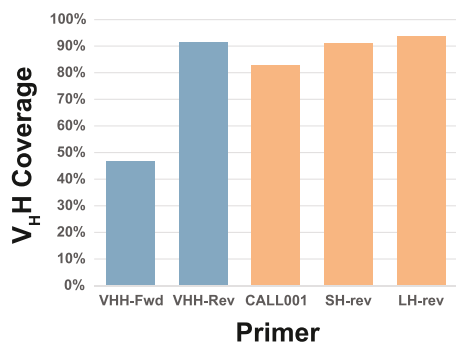
Hinge PCR:



Hinge vs. V_HH primers:



B



C

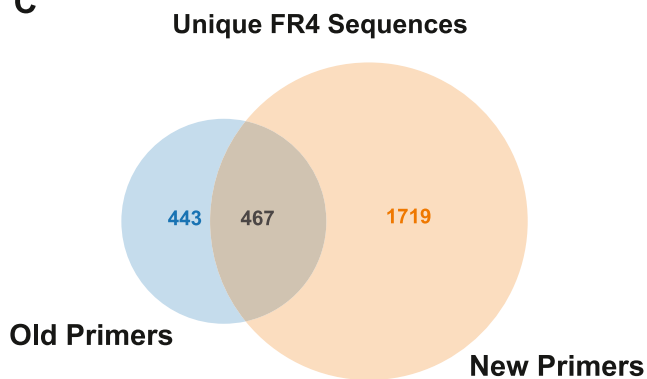


Figure 3. Amplification of V_HH fragments by high-coverage hinge-specific primers. A, schematic representation of hinge primers selected for higher coverage single-step PCR amplification of V_HH regions. Forward primer CALL001 anneals in the leader region upstream of FR1, while reverse primers LH-rev and SH-rev anneal in the long-hinge or short-hinge region, respectively, downstream of FR4. In contrast, second step nested primers VHH-Fwd and VHH-Rev anneal in FR1 or FR4 regions. B, high-throughput sequencing of lymphocyte mRNA was used to determine what percentage of V_HH -encoding transcripts could be recognized by primers used for traditional nested PCR (blue) or hinge-targeted PCR (yellow) by searching each primer against the total sequence database. C, Venn diagram comparing the number of unique FR4 sequences captured using each primer set, as determined by adjacent sequence in V_HH transcript. New hinge primers correspond to a wider diversity of FR4 sequences. Incomplete sequencing coverage, FR4 mismatches from “old” V_HH primers, and amplification of non-IgG sequences contribute to nonoverlapping sequences.

interest. GFP was chosen as a benchmark for comparison with previous studies, and remains of intense biomedical interest, with sustained demand for reagents of the highest possible affinity for a variety of applications. After GFP immunization of a single llama, we assessed the top 20 candidates from MS analysis of lymphocyte V_HH and affinity purified serum, confirming binding of 16 by an *in vitro* binding screen. Surface

plasmon resonance (SPR) was used to measure the binding kinetics and affinities for the 12 of these nanobodies that gave the best immunofluorescence microscopy signal (see below), all of which were found to bind strongly (K_D s from 2.7×10^{-8} to $<1.4 \times 10^{-13}$ M) (Table 1 and Fig. S3). These nanobodies were also tested for affinity isolations after covalent conjugation to magnetic Dynabeads. In a representative purification of

Table 1
Summary of SPR binding kinetics data

Nanobody ID	k_{on} (1/s)	k_{off} (1/MS)	K_D (M)	k_a sem	k_d sem	K_D sem	n
LaG94-1	1.1E+06	1.8E-05	1.6E-11	3.0E+05	2.3E-06	4.7E-12	5
LaG94-2	1.1E+05	3.0E-03	2.7E-08	1.4E+04	4.1E-04	4.8E-09	3
LaG94-3	1.1E+07	1.5E-03	1.3E-10	1.5E+06	2.5E-04	2.9E-11	3
LaG94-4	4.1E+05	1.5E-03	3.7E-09	4.5E+04	1.8E-04	5.9E-10	3
LaG94-5	5.6E+06	<5.0E-06	<8.9E-13	7.8E+05	n/a	1.2E-13	5
LaG94-6	5.2E+06	4.3E-04	8.3E-11	8.1E+05	5.9E-05	1.7E-11	3
LaG94-8	5.5E+06	1.2E-03	2.1E-10	4.3E+06	7.5E-04	2.1E-10	3
LaG94-9	4.5E+07	9.5E-04	2.1E-11	2.4E+07	3.7E-04	1.4E-11	3
LaG94-10	3.9E+06	3.5E-05	9.0E-12	1.1E+06	9.9E-06	3.6E-12	5
LaG94-12	1.3E+06	4.1E-04	3.0E-10	1.9E+05	5.3E-05	5.8E-11	3
LaG94-15	3.6E+07	<5.0E-06	<1.4E-13	2.2E+07	n/a	8.6E-14	3
LaG94-18	3.3E+05	3.1E-04	9.2E-10	9.2E+04	4.1E-05	2.8E-10	3
LaTdT-1	2.8E+05	<5.0E-06	<1.8E-11	6.0E+04	n/a	3.8E-12	2
LaTdT-2	1.3E+07	4.5E-05	3.5E-12	1.1E+06	1.4E-05	1.1E-12	2
LaTdT-8	1.3E+07	4.8E-04	3.7E-11	4.4E+06	1.1E-04	1.5E-11	4
LaTdT-9	2.1E+05	1.8E-05	8.4E-11	4.8E+04	1.0E-05	5.2E-11	4
LaTdT-10	7.9E+05	4.1E-05	5.2E-11	1.5E+05	1.4E-05	2.1E-11	4
LaTdT-38	1.7E+06	9.7E-05	5.7E-11	2.9E+05	1.8E-05	1.5E-11	4
LaTdT-39	1.1E+06	6.3E-06	5.9E-12	4.8E+04	1.0E-05	5.2E-11	3
LaTdT-44	1.1E+06	4.9E-05	4.4E-11	2.7E+05	1.2E-05	1.5E-11	2
LaTdT-45	4.7E+05	1.3E-05	2.8E-11	4.4E+03	2.8E-07	6.4E-13	2
GST-1	1.1E+06	9.2E-05	8.5E-11	1.2E+05	4.2E-06	1.0E-11	3
GST-2	1.2E+06	7.9E-05	6.4E-11	4.3E+04	2.1E-05	1.7E-11	2
GST-3	1.4E+06	4.8E-04	3.5E-10	6.0E+05	7.1E-05	1.6E-10	3
GST-4	4.0E+05	1.1E-04	2.6E-10	1.5E+05	3.5E-05	1.3E-10	3
GST-5	8.7E+05	1.1E-04	1.3E-10	2.3E+05	3.9E-05	5.7E-11	3
GST-6	4.1E+06	6.1E-04	1.5E-10	2.2E+05	3.8E-05	1.2E-11	2
GST-7	4.3E+06	7.6E-04	1.7E-10	7.8E+04	6.6E-05	1.6E-11	2
GST-8	2.4E+05	3.1E-05	1.3E-10	9.8E+04	1.0E-05	6.9E-11	2
LaMlgG-5	2.8E+05	1.2E-04	4.2E-10	1.4E+05	3.6E-06	2.1E-10	2
LaMlgG-8	1.3E+05	1.6E-04	1.2E-09	2.8E+04	1.7E-06	2.8E-10	2
LaMlgG-9	2.9E+05	1.8E-04	6.0E-10	3.3E+04	1.9E-05	9.3E-11	2
LaMlgG-14	2.2E+05	4.7E-05	2.1E-10	1.6E+04	8.6E-06	4.3E-11	2
LaRIgG-1 ^a	8.9E+05	3.4E-03	3.6E-10	1.4E+04	7.1E-04	9.6E-11	2
	2.6E-03	2.6E-04		1.5E-04	2.9E-05		
LaRIgG-2	3.6E+05	<5.0E-06	<1.4E-11	1.5E+05	n/a	6.0E-12	4
LaRIgG-7	9.6E+05	5.1E-05	5.3E-11	1.2E+05	1.0E-06	6.6E-12	2
LaGIgG-4 ^a	2.2E+06	1.1E-02	8.2E-10	1.9E+05	1.9E-03	3.6E-10	2
	1.8E-03	3.3E-04		2.7E-04	3.0E-05		
LaGIgG-12	3.2E+05	1.6E-04	4.9E-10	6.7E+04	2.6E-05	1.3E-10	3

^a 2-state binding model.

Nup84-GFP from yeast, the tested nanobodies isolated the associated Nup84 complex at comparable or better yield and purity to the best previously identified GFP nanobody dimer (Fig. 4A) (14).

Alongside GFP, we coimmunized the same animal with tdTomato, a highly divergent synthetic red fluorescent protein dimer derived from *Discosoma* sp (25). This is a widely used tag often applied in transgenic animals for immunohistochemistry (IHC) studies due to its favorable brightness, maturation, and folding properties (26, 27) and for which a high-affinity bait is desirable, e.g., for complementary interactomic studies. We again identified 31 top candidates, 21 of which passed our initial binding screen. Nine were assessed by SPR, demonstrating consistently excellent affinity (K_D s from 8.4×10^{-11} to 3.5×10^{-12} M) (Table 1 and Fig. S4).

As an additional target we selected GST, most commonly used as a fusion protein tag for recombinant protein expression and affinity purification (28). While immobilized glutathione is routinely used for GST isolation, anti-GST nanobodies provide complementary tools for isolation or detection of GST-tagged proteins, enabling new applications due to their flexibility as encodable, modifiable proteins (29). We generated eight high-affinity nanobodies against GST, verified by SPR (K_D s from 3.5×10^{-10} to 6.4×10^{-11} M)

(Table 1 and Fig. S5). As a benchmark for specificity and affinity, we compared these nanobodies to glutathione Sepharose in affinity isolations of a GST-tagged protein (GST-tobacco etch virus [TEV]) from *Escherichia coli* lysate. In increasing dilutions of the tagged protein, at least two nanobodies showed superior recovery of GST, with comparable high purity (Fig. 4B). With this high affinity, high specificity performance, these reagents are well suited to the efficient capture or detection of GST-tagged targets.

We further immunized three animals (one alpaca and two llamas) with IgG from mouse, rabbit, or goat. Given the ubiquity of antibodies generated in these species, new anti-IgG nanobodies continue to be valuable tools for use as single domain substitutes for traditional secondary antibodies. Our pipeline identified 14, 6, and 12 nanobodies against mouse, rabbit, and goat IgG respectively, from 16, 9, and 15 top candidates. SPR again demonstrated high affinities for nine tested IgG nanobodies (K_D from 1.2×10^{-9} M to $<1.4 \times 10^{-11}$) (Table 1 and Fig. S6).

Nanobodies as optimized tools for immunofluorescence microscopy

Sensitive fluorescent detection of protein tags, particularly IHC in fixed cells or tissues, demands high-performance

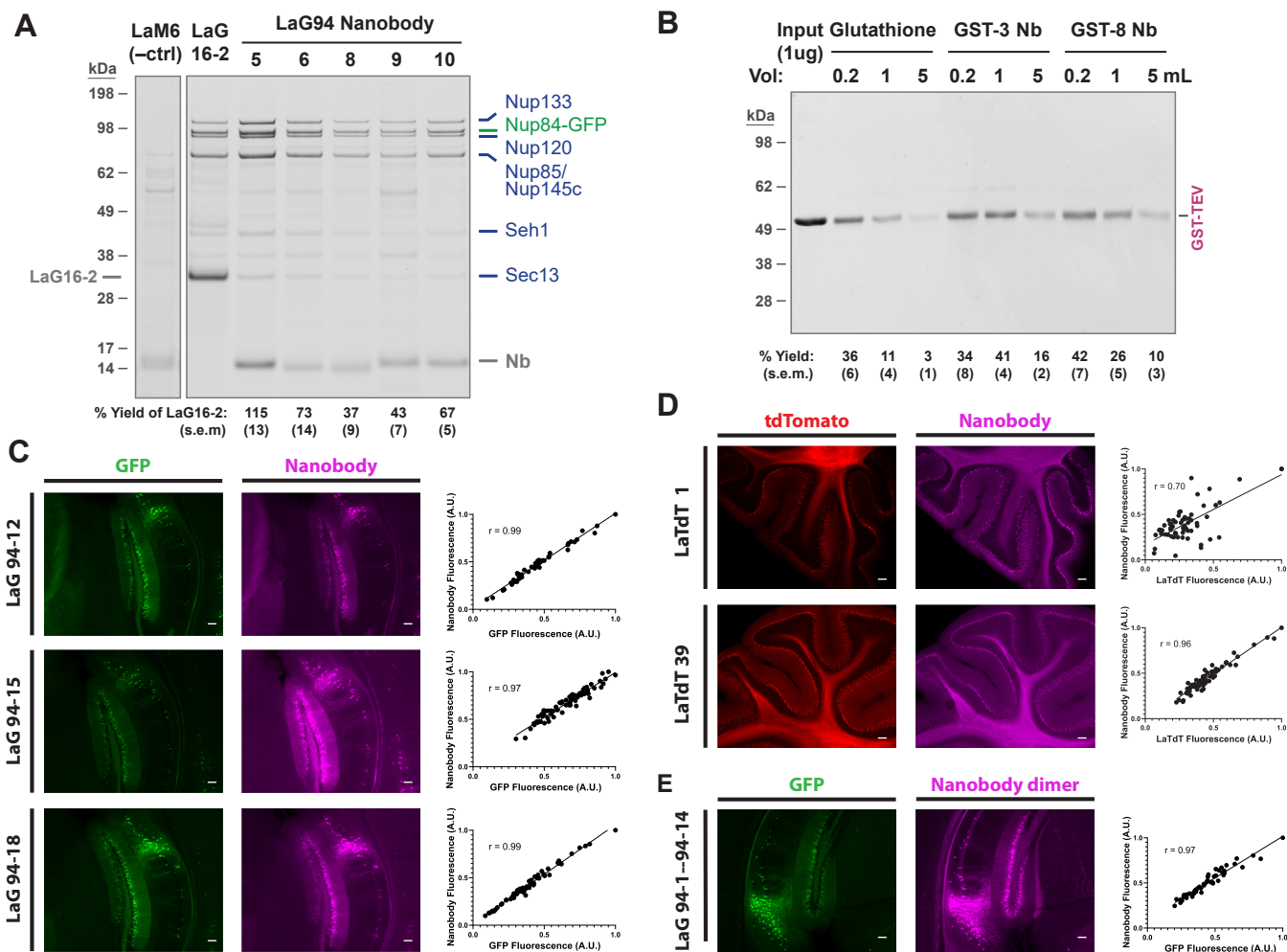


Figure 4. GFP and tdTomato nanobody performance in tissue IF and affinity purifications. A, GFP nanobodies were conjugated to Dynabeads and used to affinity purify Nup84-GFP and its associated NPC subcomplex from yeast. An anti-mCherry nanobody (LaM6) was used as a negative control for nonspecific background, and the LaG16-2 nanobody heterodimer used as the highest affinity nanobody reagent previously reported (14). Complexes were eluted in SDS and analyzed by Coomassie-stained SDS-PAGE. Band densities of Nup133, Nup84-GFP, Nup120, and Nup85/Nup145c were quantified and normalized to the LaG16-2 lane (n = 3). B, Briefly, 1 μ g of GST-TEV was diluted in 0.2 to 5 ml of *Escherichia coli* lysate and isolated with glutathione- or nanobody-coupled Sepharose, with SDS elutions analyzed by Coomassie-stained SDS-PAGE. Recovery of the GST-tagged protein was quantified by densitometry (n = 3). C, new GFP nanobodies were labeled with Alexa Fluor 568 and used to stain brain sections of Thy1-GFP mice. Quantitative comparisons of relative GFP and nanobody fluorescence are plotted at right. Fluorescence was normalized to maximum fluorescence in each channel. Plots were fitted with simple linear regression, and Pearson correlation coefficients (r) are indicated. The scale bars represent 100 μ m. D, tdTomato nanobodies were labeled with Alexa Fluor 647 and used to stain sagittal brain sections of ChAT-Cre::Ai14 mice, with tdTomato expressed in cholinergic neurons. Quantification is shown as in (C). The scale bars represent 100 μ m. E, a GFP nanobody heterodimer was labeled with Alexa Fluor 647 and used for Thy1-GFP brain section staining as in (C). The scale bars represent 100 μ m. IF, immunofluorescence; NPC, nuclear pore complex; TEV, tobacco etch virus.

antibodies. Dye-labeled nanobodies offer particular advantages in this application. Their small size allows significantly faster penetration into samples, particularly important for volume imaging where antibody diffusion is a significant bottleneck (30, 31). This small size also contributes to more precise localization in high-resolution microscopy. Further, nanobodies bind at a single epitope, and thus potentially can have a 1:1 binding ratio that leads to a linear response in signal, an important factor in quantitative analysis.

To evaluate the effectiveness of our new nanobody repertoires in IHC, we screened fluorescently labeled nanobodies against brain sections from transgenic mice expressing GFP or tdTomato in the brain (Thy1-GFP or ChAT-Cre/Ai14 tdTomato reporter). Out of 16 GFP nanobodies tested, nine demonstrated the strongest fluorescence correlating to GFP

signal (Figs. 4C and S7). Similarly, five tdTomato nanobodies proved the most effective out of 20 screened (Figs. 4D and S8). Importantly, the fluorescent signal intensity of these GFP nanobodies has a strong linear correlation to GFP itself (Pearson $r \geq 0.92$). While this correlation, and overall signal to noise performance, is somewhat weaker for most tdTomato candidates, LaTdT-39 performs at a similarly high level ($r = 0.96$).

Epitope screening and oligomerization of GFP nanobodies for immunohistochemistry

To further improve the sensitivity and specificity of nanobodies in this application, we sought to engineer bivalent, heterodimeric constructs with increased avidity. To determine

the optimal combinations of nanobodies for this purpose, we first binned epitopes by complementary staining of nanobody pairs in IHC. We have previously established the binding sites of reported GFP nanobodies, which fell into three nonoverlapping epitope groups (14). Representative nanobodies binding these three epitopes (LaG16, LaG41, and LaG2) were used to screen new GFP nanobodies by competitive binding in IF. These controls were labeled with a different fluorophore, and used to pre-stain a GFP tissue sample to saturation before probing with the new nanobody set (Figs. S9–S11) (14). When we observed simultaneous fluorescent signal from both nanobodies, the pair was judged to be complementary, binding distinct epitopes. In addition to providing biochemical epitope information, this approach revealed which pairs of nanobodies could achieve complementary, simultaneous binding in the IHC environment of practical interest. Most of the tested anti-GFP nanobodies did in fact show complementation with the three representative nanobodies tested. Three (LaG94-1, LaG94-10, and LaG94-12) appear to overlap with LaG41's epitope, and in fact outcompete LaG41 during costaining. To further classify the remaining nanobodies, we performed a similar complementation screen between LaG94-10, 14, 15, and 18 (Fig. S12). All four were able to complement each other, indicating distinct epitopes.

With the numerous complementary nanobody pairs revealed by these experiments, we generated a series of dimeric constructs. These were again assessed in IHC of Thy1-GFP brain sections, and while more variance was found in signal to noise and linearity performance, at least two dimers, LaG94-1—LaG94-14 and LaG94-14—LaG94-12, showed significantly improved specificity and signal strength while retaining linear staining (Figs. 4E and S13).

Anti-IgG nanobodies as improved secondary antibodies

Mouse, rabbit, and goat IgG nanobodies were similarly tested against Thy1-GFP brain sections first stained with anti-GFP monoclonal antibodies from the corresponding species (Figs. 5 and S14). Strong, linear staining was found with several of these nanobodies: 5 anti-mouse, 3 anti-rabbit, and 1 anti-goat. The signal specificity for these top performers was comparable to a standard mouse and goat polyclonal secondary antibody; however, the linear correlation of signal was far higher (Pearson $r = 0.76$ – 0.98 versus $r = -0.03$ – 0.15). These monoclonal nanobody agents are thus significantly better suited to IHC applications where antigen localization must be quantitatively compared.

We further categorized the species and isotype specificity of the IgG nanobodies effective in IHC by immunoblotting (Fig. S15). Most were highly specific for IgG from their target species, with the exceptions of goat G-4 and G-12 (sheep cross-reactivity), rabbit R-2 (human cross-reactivity), and mouse M-5, M-8, and M-14 (rat cross-reactivity). Mouse nanobody M-9 bound all mouse IgG isotypes other than IgG1 λ , while others specifically targeted IgG2b. The tested nanobodies targeted the Fc domain, other than rabbit R-1 and the mouse nanobodies, which bound F(ab')₂.

Discussion

For any nanobody application, rapid and efficient identification of high-affinity clones is a critical first step. We previously demonstrated a nanobody pipeline making use of MS and next-generation DNA sequencing, with advantages in rapid generation of large repertoires, and identification of ultra-high-affinity candidates from high-stringency selection. We have now made a variety of improvements to this pipeline, significantly increasing coverage of V_HH libraries from immunized animals and improving the quality of MS samples, including in the cases of highly conserved or weakly reactive immunogens. This pipeline can generate large numbers of nanobody candidates as needed, and we have for example validated well over 100 high-affinity nanobodies from a single antigen (4).

Using these improved methods, we have identified new repertoires of nanobodies against a panel of protein antigens with immediate value in research applications: GFP, tdTomato, GST, and rabbit, mouse, and goat IgG. These repertoires build significantly on previously reported nanobodies against these or related targets (14, 29, 32, 33), *i.e.*, with significantly higher affinity, increased repertoire size covering multiple epitopes, and against new and widely used antigens like tdTomato and goat IgG. These nanobodies have affinities into the picomolar range, and demonstrate strong activity in both affinity capture and IF applications. We have increased affinity and performance even further by identifying epitope bins and engineering dimeric nanobodies, focusing on anti-GFP reagents. Nanobodies have a variety of intrinsic advantages as affinity reagents, particularly in IF applications. The 10-fold smaller size of nanobodies compared to standard IgG secondaries provides for faster diffusion, allowing for more efficient staining of thicker sections or whole-mount samples (30, 31). This small size is a further advantage for super-resolution microscopy, with the significantly closer proximity of the covalently conjugated fluorophore allowing more precise localization of the target antigen (34, 35).

Experimental procedures

Llama immunizations and bleeds

Llama immunizations were performed as previously described (14). Briefly, naïve animals were immunized with an antigen mixture containing 200 μ g to 1 mg of each antigen protein, emulsified with one volume of complete Freund's adjuvant. Three subsequent boosts were performed at 21 days intervals, using incomplete Freund's adjuvant. Animals previously immunized (>6 months prior) instead received both primary and boost immunizations with incomplete Freund's adjuvant. Production serum bleeds and bone marrow aspirates were obtained 10 days after the final boost.

V_HH sequencing

Lymphocytes were isolated from bone marrow aspirates using a Ficoll-Paque gradient separation (Cytiva). RNA was isolated using TRIzol (Thermo Fisher Scientific), and cDNA

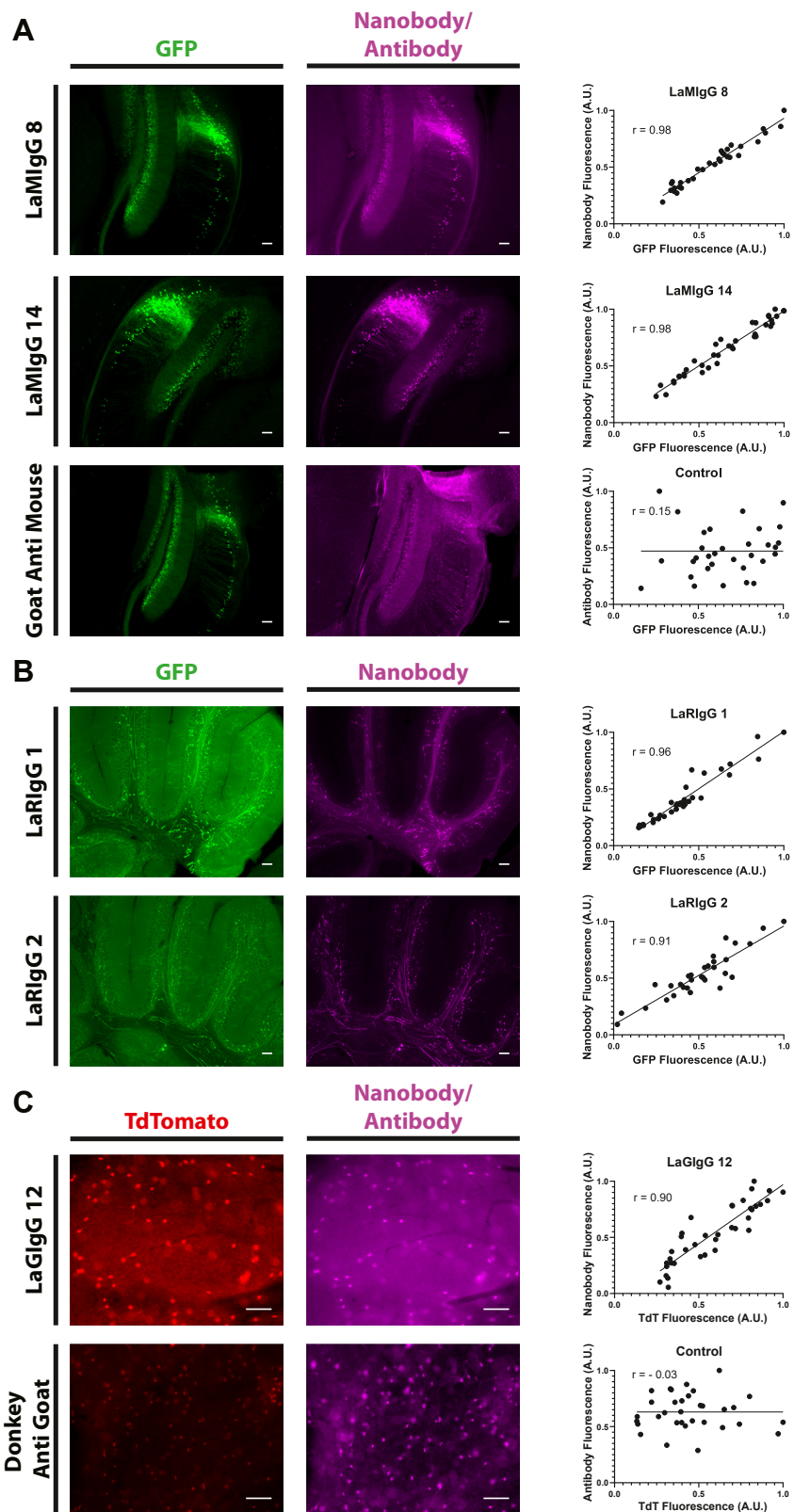


Figure 5. IgG nanobodies as linear secondary detectors in tissue IF. Nanobodies against mouse IgG (A), rabbit IgG (B), or goat IgG (C) were fluorescently labeled with Alexa Fluor 568 (mouse and rabbit) or Alexa Fluor 647 (goat) and used to detect anti-GFP or anti-tdTomato primary antibodies in Thy1-GFPM (A and B) or somatostatin-TdTomato (C) mouse brain sections. Standard monoclonal anti-mouse and polyclonal anti-goat secondaries were also tested. Quantitative comparisons of relative GFP or tdTomato and nanobody/secondary fluorescence are plotted at *right*. Fluorescence was normalized to maximum fluorescence in each channel. Plots were fitted with simple linear regression, and Pearson correlation coefficients (r) are indicated. The scale bars represent 100 μm . IF, immunofluorescence; IgG, immunoglobulin G.

was generated using SuperScript IV reverse transcriptase with oligo(dT) primers (Thermo Fisher Scientific). V_HH domains were then PCR amplified using Deep Vent polymerase (NEB). Nested PCRs were performed as previously described using CALL001/CALL002 primers, followed by MiSeq-VHH-forward and MiSeq-VHH-reverse primers (14). Single-step PCRs using hinge-specific primers were performed with 6N-CALL001/6N-CALL001b (5'-NNNNNN GTCCTGGCT GCTCTT[C/T]TACAAGG-3') forward primers and a 1:1 mixture of 6N-SH-rev (5'-NNNNNN CTGGGGTCTTCGC TGTGGTGC-3') and 6N-LH-rev (5'-NNNNNN GTGTTG TGGTTTGGTGTCTTGGG-3') reverse primers. PCR products were gel-purified, ligated to Illumina adapters for library preparation, and sequenced by MiSeq with 2 × 300-bp paired end reads.

CALL001 and CALL002 primers were used to PCR amplify IgG fragments from llama and alpaca bone marrow lymphocyte cDNA, which were gel-purified and prepared using a 20kb SMRTBell library prep kit (PacBio) for sequencing on a PacBio RS II system. For sequencing of llama lymphocyte transcripts, mRNA was enriched from total RNA by oligo(dT) capture, and libraries prepared using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina). NextSeq Mid Output 2 × 75-bp paired end sequencing was carried out on an Illumina NextSeq 500 instrument.

For *in silico* protease digests, sequences from translated databases (see Llama-Magic 2.0 below) were numbered using ANARCI software with IMGT numbering (<https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabpred/anarci/>) (36). *In silico* digestion was performed using Pyteomics with default Expsy cleavage rules for trypsin, chymotrypsin high specificity, and chymotrypsin low specificity, and no missed cleavages (https://pyteomics.readthedocs.io/en/latest/_modules/pyteomics/parser.html) (37, 38).

HCAb preparation

Heavy-chain only variant IgG (HCAb) was first enriched from crude serum using Protein A and Protein G as previously described (1, 14). Protein M was separately expressed and purified in bacteria, and conjugated to CNBr-activated Sepharose 4 Fast Flow (Cytiva) at a ratio of 1 mg Protein M to 50 mg Sepharose (final coupling density of approximately 5 mg per ml of swelled Sepharose). Residual light-chain containing IgG was then depleted from Protein A/G-fractionated HCAb by incubating with PrM-Sepharose (50 µl Sepharose bed volume per 1 mg IgG) for 30 min at room temperature and collecting flow-through.

HCAb affinity capture and IdeS digestion

In total, 10 mg of purified HCAb was affinity purified over 100 µl of CNBr-Sepharose coupled to the corresponding antigen protein at 1 mg/ml. The resin was washed sequentially with PBS-500 (500 mM NaCl final concentration), PBS + 0.5% Triton, 2 to 3 M MgCl₂/10 mM Tris pH 7.5, and PBS. The resin was then resuspended in 200 µl PBS and digested with 400 units of IdeS (FabRICATOR, Genovis) for 3.5 h, mixing at

37 °C. Comparative samples were digested with papain as previously described (14), or eluted in lithium dodecyl sulfate (LDS) without digestion. Bound protein was eluted by incubating 10 min at 72 °C with 1.2× LDS loading buffer (Thermo Fisher Scientific). Samples were reduced, alkylated with iodoacetamide, and separated by SDS-PAGE. V_HH bands at 15 to 25 kDa were excised and prepared for MS analysis.

Mass spectrometry

Trypsin (Promega) or chymotrypsin (Roche) solution was added to diced, destained, and dehydrated gel pieces at ~1:4 enzyme to substrate mass ratios in 55 µl digestion buffer (trypsin: 50 mM ammonium bicarbonate; chymotrypsin: 100 mM Tris pH 8.0, and 2 mM CaCl₂). In the case of trypsin, enzyme was supplied in two additions of equal mass each. After the first addition, the samples were incubated at 37 °C for 4 h. After the second addition, the samples were incubated at 37 °C overnight. For chymotrypsin, one of the incubation was performed for 1.5 h at 37 °C. Supernatants were removed from gel pieces and transferred to new tubes. To extract peptides, 100 µl of a 1.7% v/v formic acid, 70% v/v acetonitrile, and 0.1% v/v trifluoroacetic acid solution were added to the gel pieces, and tubes were shaken at 37 °C for 1 h. Supernatant was removed from gel pieces and combined with previous supernatant. Sequential extractions were performed similarly with 100 mM triethylammonium bicarbonate and acetonitrile. The pooled supernatants were evaporated in a speedvac until dry. Peptides were resuspended in 5% v/v formic acid, 0.1% v/v trifluoroacetic acid, and cleaned on StageTips (39).

Samples were analyzed with a nano-LC 1200 (Thermo Fisher Scientific) using an EASyspray PepMap RSLC C18 3 micron, 100 Å, 75 µm by 15 cm column coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). The instrument was operated in data-dependent mode, and top intensity ions were fragmented by collision-induced dissociation (collision energy 35). Ions with charge states 2 to 6 were selected for fragmentation. The quadrupole isolation window was 1.4, and the MS/MS used a maximum injection time of 250 ms with 1 microscan.

Llama-Magic 2.0

The Llama-Magic github repository is located at https://github.com/FenyoLab/Llama_Nanobodies. It includes all code, as well as detailed instructions for running the custom scripts for database preparation and for setting up and using the Llama-Magic web tool. Detailed methods are included in supporting information.

In brief, to identify candidate nanobody sequences, a custom database of *in silico* translated protein sequences is created from the DNA sequencing reads. After merging the read pairs, the correct ORF is selected by determining the position of the PCR primers near the start and end of each sequence. The protein sequence database is then *in silico* digested with both trypsin and chymotrypsin and lists of unique tryptic and chymotryptic peptides are saved as FASTA files. The tandem MS data are searched against these peptide lists and identified

Nanobody research tools using improved MS-based discovery

peptides from both tryptic and chymotryptic digests are mapped back to the protein sequences. In order to determine MS coverage over the CDR regions, we implement a custom algorithm that searches for specific conserved amino acid motifs proximate to CDR start and end sites. Each sequence is scored and ranked based on coverage over the CDR regions, with CDR3 coverage carrying the most weight. Sequences with similar CDR3 regions are grouped together in the output. Each identified peptide is given a uniqueness score that assesses its uniqueness within the CDR3 group relative to how often it is found in the overall sequence database.

Protein preparation

GFP, tdTomato, and GST-TEV protease were produced in bacteria according to standard protocols. IgG proteins were purchased from Jackson Immunoresearch. Nanobodies were expressed in ArcticExpress (DE3) cells (Agilent Technologies) with a pelB leader sequence and C-term 6xHis tag, and purified from periplasm as previously described (14). For IF experiments, nanobodies were labeled with amine-reactive dyes as follows. First, 1.2 μ l of 1 M NaHCO₃ was added to 10 μ l of 1.25 mg/ml nanobody. For Alexa Fluor 568 conjugation, 1 μ l of 6 mg/ml Alexa Fluor 568 NHS Ester (Thermo Fisher Scientific, A20003) was added. For Alexa Fluor 647 conjugation 1 μ l of 8 mg/ml Alexa Fluor 647 NHS Ester (Thermo Fisher Scientific, A20006) was added. Samples were incubated at 37 °C for 1 h. Reaction was quenched by addition of PBS with 0.2% Triton X-100 and 10 μ g/ml heparin (PTxH) by diluting to 10 μ g/ml.

Binding kinetics measurements by SPR

GFP nanobody SPR was performed on a ProteOn XPR36 (Bio-Rad) or Biacore 8K (Cytiva) instrument, in both cases essentially as previously described (4, 14). The Biacore 8K was used for all other antigens. For ProteOn measurements, GFP was immobilized on a GLC sensor chip (Bio-Rad) at 5 μ g/ml in 10 mM sodium acetate pH 5.0 after EDC/NHS activation. Parallel binding measurements were performed at four or five nanobody concentrations run at 50 μ l/min, with an association time of 180 s and a dissociation time of 960 s. Data processing and analysis were done with ProteOn Manager software (<https://www.bio-rad.com/en-us/product/proteon-manager-software>), fitting to a Langmuir binding model.

For Biacore measurements, antigens were immobilized at 1 μ g/ml (tdTomato) or 5 μ g/ml (GFP, GST, and IgGs) in 10 mM sodium acetate pH 5.5 on Series S CM5 sensor chips, using an EDC/NHS amine coupling kit (Cytiva) according to the manufacturer's guidelines. Nanobodies were injected at 30 μ l/min in single-cycle kinetics experiments at concentrations of 0.1, 0.3, 1, 3, and 10 nM, with an association time of 180 s, and a dissociation time of 1800 to 7200 s, depending on observed off-rate. Data were analyzed using Biacore software (<https://www.cytivalifesciences.com/en/us/shop/protein-analysis/spr-label-free-analysis/spr-software-and-extensions/biacore-insight-evaluation-software-p-23528>), fitting a Langmuir 1:1 binding model to sensorgrams to calculate kinetic parameters.

RlgG-1 and GlgG-4 nanobodies were instead fit to a two-step binding model.

All binding experiments were conducted in a buffer of 20 mM Na-Hepes pH 7.4, 150 mM NaCl, 0.05% Tween 20. Sensor surfaces were regenerated between measurements with 10 mM glycine-HCl pH 3.0 (IgGs) or 3.5 M MgCl₂ in 10 mM Tris-HCl pH 7.5.

Affinity purifications

For yeast Nup84-GFP isolations, recombinant nanobodies were conjugated to epoxy-activated M-270 magnetic Dynabeads (Thermo Fisher Scientific) as previously described (14). Briefly, 10 μ g nanobody protein was used per 1 mg of Dynabeads, with conjugations carried out in 0.1 M sodium phosphate, pH 8.0 and 1 M ammonium sulfate, incubating for 18 to 20 h rotating at 30 °C. Affinity isolations were carried out as previously described, using binding buffer consisting of 20 mM Hepes, pH 7.4, 500 mM NaCl, 2 mM MgCl₂, 0.1% CHAPS, 0.1 M PMSE, and 3 μ g/ml pepstatin A (14, 40). For each experiment, 25 μ l of bead slurry was used with 0.25 g of yeast cell powder resuspended in 2.25 ml buffer.

For GST isolations, nanobodies were conjugated to NHS-activated Sepharose (Cytiva) according to the manufacturer's instructions. Recombinant GST-fused TEV protease was diluted in 0.5 mg/ml lysate prepared from BL21 DE3 bacterial cells, which were resuspended in 20 mM sodium phosphate, pH 7.4, 350 mM NaCl, 0.1% Tween 20, 0.1 M PMSE, 3 μ g/ml pepstatin A, and lysed by microfluidizer. A 10 μ l bed volume of nanobody-Sepharose, or commercial glutathione Sepharose (GenScript) was incubated for 30 min with each sample, and then washed twice with 20 mM sodium phosphate, pH 7.4, 350 mM NaCl, and 0.1% Tween 20. Bound protein was eluted with LDS at 72 °C and analyzed by Coomassie-stained SDS-PAGE.

Immunoblots

For nanobody dot blots, 500 ng IgG (250 ng for subtype-specific IgG) was spotted in 1 μ l volume on nitrocellulose strips. After drying, membranes were blocked with 5% milk/PBS for 30 min, and then probed with Alexa Fluor 488-labeled nanobodies at 300 nM in 5% milk/PBS for 30 min. After two 10 min washes in PBS, blots were imaged with a ImageQuant LAS 4000 fluorescent imager using Cy3 filters (Cytiva).

Immunohistochemistry of brain sections

Adult female mice were anesthetized with isoflurane and fixed with intracardiac perfusion of 4% paraformaldehyde in PBS. Samples were postfixed in 4% paraformaldehyde in PBS at 4 °C overnight. Subsequently, 100 μ m sagittal sections were cut by vibratome from agar embedded brains.

Sections were washed twice with PTxH for 5 min and then 30 min. Nanobody was added at a concentration of 1 μ g/ml for 5 h. Samples were then washed with PTxH for 1, 5, and 30 min followed by two 5-min washes in PBS. Samples were then mounted. For complementation experiments, following the 1, 5, and 30 min PTxH washes, the second nanobody was added

at a concentration of 1 µg/ml for 5 h. Samples were then washed again in PTxH for 1, 5, and 30 min followed by two 5 min washes in PBS and mounting. For experiments using a nanobody secondary, primary antibodies were left on the section overnight. For LaMIgGs, a mouse anti-GFP (19C8, Memorial Sloan Kettering Monoclonal Antibody Facility) (41) at 1:500 was used as the primary antibody; for LaRIgGs, a polyclonal rabbit anti-GFP (abcam, ab290) at 1:500 was used; for the LaGIgGs, a polyclonal goat anti-mCherry (Scigen, Cat # AB0081-500) at 1:1000 was used. For LaMIgG the control secondary was 1:500 Goat anti-Mouse conjugated to Alexa Fluor 568 (Invitrogen). For LaGIgG control secondary was 1:500 Donkey anti Goat conjugated to Alexa Fluor 647 (Invitrogen). LaMIgG, LaGIgG, LaRIgG, and control secondaries were left on sections for 4 h.

Imaging and image analysis

Images were acquired on a Nikon eclipse 90i upright fluorescent microscope, using either a 4× or 10× objective. Images were processed in ImageJ (<https://imagej.net/>). For Pearson's Analysis the fluorescence of at least 40 cell somas was quantified. For each section, the values were normalized to the maximum fluorescence value in that channel. Values were plotted in GraphPad Prism (<https://www.graphpad.com/>) to calculate the Pearson correlation coefficient (*r*).

Animal husbandry

All mouse care was performed in compliance with the protocols approved by the Institutional Care and Use Committee (IACUC) of The Rockefeller University. Llama care was performed at Capralogics, Inc according to protocols approved by their IACUC. The Thy1-GFPM, Ai14 (Rosa26^{lox-stop-lox-tdTomato}), and ChAT-cre lines were acquired from the Jackson Laboratory.

Data availability

Mass spectrometry datasets are deposited in the Zenodo repository, <https://doi.org/10.5281/zenodo.10805358>. Nanobody sequences are provided in supporting information (Table S1).

Supporting information—This article contains supporting information (42–44).

Acknowledgments—We thank the Fisher Drug Discovery Resource Center at Rockefeller University (RRID:SCR_020985) for SPR assistance. We thank Yi Shi for helpful discussions on mass spectrometric methods, and Marc Tessier-Lavigne for support in study design.

Author contributions—P. C. F., R. J. F., K. R. M., E. Y. J., Y. L., J. T., and V. S. investigation; P. C. F., R. J. F., K. R. M., J. W., E. Y. J., Y. L., D. F., Z. W., B. T. C., and M. P. R. methodology; P. C. F., R. J. F., K. R. M., E. Y. J., Y. L., D. F., Z. W., B. T. C., and M. P. R. writing—review and editing; P. C. F., R. J. F., K. R. M., S. K. and E. Y. J., writing—original draft; P. C. F., R. J. F., Z. W., B. T. C., and M. P. R.

conceptualization; S. K., J. W., and D. F. software; D. F., B. T. C., and M. P. R. supervision; B. T. C. and M. P. R. funding acquisition.

Funding and additional information—This work was supported by the National Institutes of Health grant P41GM109824 (B. T. C. and M. P. R.), the G. Harold and Leila Y. Mathers Charitable Foundation (B. T. C. and M. P. R.), the Robertson Therapeutic Development Fund (B. T. C. and M. P. R.), and the Jain Foundation (B. T. C. and M. P. R.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: cDNA, complementary DNA; CDR3, complementarity determining region 3; IHC, immunohistochemistry; LDS, lithium dodecyl sulfate; MS, mass spectrometry; SPR, surface plasmon resonance; TEV, tobacco etch virus.

References

- Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., *et al.* (1993) Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448
- Muyldermans, S. (2013) Nanobodies: natural single-domain antibodies. *Annu. Rev. Biochem.* **82**, 775–797
- Jovcevska, I., and Muyldermans, S. (2020) The therapeutic potential of nanobodies. *BioDrugs* **34**, 11–26
- Mast, F. D., Fridy, P. C., Ketaren, N. E., Wang, J., Jacobs, E. Y., Olivier, J. P., *et al.* (2021) Highly synergistic combinations of nanobodies that target SARS-CoV-2 and are resistant to escape. *Elife* **10**, e73027
- Dumoulin, M., Conrath, K., Van Meirhaeghe, A., Meersman, F., Heremans, K., Frenken, L. G., *et al.* (2002) Single-domain antibody fragments with high conformational stability. *Protein Sci.* **11**, 500–515
- Chanier, T., and Chames, P. (2019) Nanobody engineering: toward next generation immunotherapies and immunomaging of cancer. *Antibodies (Basel)* **8**, 13
- Duggan, S. (2018) Caplacizumab: first global approval. *Drugs* **78**, 1639–1642
- Yong Joon Kim, J., Sang, Z., Xiang, Y., Shen, Z., and Shi, Y. (2023) Nanobodies: robust miniprotein binders in biomedicine. *Adv. Drug Deliv. Rev.* **195**, 114726
- Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R., and Muyldermans, S. (1997) Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett.* **414**, 521–526
- Pardon, E., Laeremans, T., Triest, S., Rasmussen, S. G., Wohlkonig, A., Ruf, A., *et al.* (2014) A general protocol for the generation of Nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693
- McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., *et al.* (2018) Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol.* **25**, 289–296
- Uchanski, T., Zogg, T., Yin, J., Yuan, D., Wohlkonig, A., Fischer, B., *et al.* (2019) An improved yeast surface display platform for the screening of nanobody immune libraries. *Sci. Rep.* **9**, 382
- Cross, F. R., Fridy, P. C., Ketaren, N. E., Mast, F. D., Li, S., Olivier, J. P., *et al.* (2023) Expanding and improving nanobody repertoires using a yeast display method: targeting SARS-CoV-2. *J. Biol. Chem.* **299**, 102954
- Fridy, P. C., Li, Y., Keegan, S., Thompson, M. K., Nudelman, I., Scheid, J. F., *et al.* (2014) A robust pipeline for rapid production of versatile nanobody repertoires. *Nat. Methods* **11**, 1253–1260
- Xiang, Y., Sang, Z., Bitton, L., Xu, J., Liu, Y., Schneidman-Duhovny, D., *et al.* (2021) Integrative proteomics identifies thousands of distinct, multi-epitope, and high-affinity nanobodies. *Cell Syst.* **12**, 220–234.e9

Nanobody research tools using improved MS-based discovery

- Shen, Z., Xiang, Y., Vergara, S., Chen, A., Xiao, Z., Santiago, U., *et al.* (2021) A resource of high-quality and versatile nanobodies for drug delivery. *iScience* **24**, 103014
- Grover, R. K., Zhu, X., Nieuwsma, T., Jones, T., Boreo, I., MacLeod, A. S., *et al.* (2014) A structurally distinct human mycoplasma protein that generically blocks antigen-antibody union. *Science* **343**, 656–661
- von Pawel-Rammigen, U., Johansson, B. P., and Bjorck, L. (2002) IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *EMBO J.* **21**, 1607–1615
- Bond, C. J., Marsters, J. C., and Sidhu, S. S. (2003) Contributions of CDR3 to V H H domain stability and the design of monobody scaffolds for naive antibody libraries. *J. Mol. Biol.* **332**, 643–655
- Gilar, M., Olivova, P., Chakraborty, A. B., Jaworski, A., Geromanos, S. J., and Gebler, J. C. (2009) Comparison of 1-D and 2-D LC MS/MS methods for proteomic analysis of human serum. *Electrophoresis* **30**, 1157–1167
- Conrath, K. E., Lauwereys, M., Galleni, M., Matagne, A., Frere, J. M., Kinne, J., *et al.* (2001) Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the camelidae. *Antimicrob. Agents Chemother.* **45**, 2807–2812
- Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., *et al.* (2006) Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat. Methods* **3**, 887–889
- Achour, I., Cavalier, P., Tichit, M., Bouchier, C., Lafaye, P., and Rougeon, F. (2008) Tetrameric and homodimeric camelid IgGs originate from the same IgH locus. *J. Immunol.* **181**, 2001–2009
- Kastelic, D., Frkovic-Grazio, S., Baty, D., Truan, G., Komel, R., and Pompon, D. (2009) A single-step procedure of recombinant library construction for the selection of efficiently produced llama VH binders directed against cancer markers. *J. Immunol. Methods* **350**, 54–62
- Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572
- Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007) A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605
- Morris, L. M., Klanke, C. A., Lang, S. A., Lim, F. Y., and Crombleholme, T. M. (2010) TdTomato and EGFP identification in histological sections: insight and alternatives. *Biotech. Histochem.* **85**, 379–387
- Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31–40
- Groll, N., Emele, F., Poetz, O., and Rothbauer, U. (2015) Towards multiplexed protein-protein interaction analysis using protein tag-specific nanobodies. *J. Proteomics* **127**, 289–299
- Cai, R., Pan, C., Ghasemigharagoz, A., Todorov, M. I., Forstera, B., Zhao, S., *et al.* (2019) Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull-meninges connections. *Nat. Neurosci.* **22**, 317–327
- Li, J., Czajkowsky, D. M., Li, X., and Shao, Z. (2015) Fast immuno-labeling by electrophoretically driven infiltration for intact tissue imaging. *Sci. Rep.* **5**, 10640
- Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., *et al.* (2010) Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* **17**, 133–138
- Pleiner, T., Bates, M., and Gorlich, D. (2018) A toolbox of anti-mouse and anti-rabbit IgG secondary nanobodies. *J. Cell Biol.* **217**, 1143–1154
- Ries, J., Kaplan, C., Platonova, E., Eghlidi, H., and Ewers, H. (2012) A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat. Methods* **9**, 582–584
- Carrington, G., Tomlinson, D., and Peckham, M. (2019) Exploiting nanobodies and Affimers for superresolution imaging in light microscopy. *Mol. Biol. Cell* **30**, 2737–2740
- Dunbar, J., and Deane, C. M. (2016) ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* **32**, 298–300
- Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V., and Gorshkov, M. V. (2013) Pyteomics—a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.* **24**, 301–304
- Levitsky, L. I., Klein, J. A., Ivanov, M. V., and Gorshkov, M. V. (2019) Pyteomics 4.0: five years of development of a Python proteomics framework. *J. Proteome Res.* **18**, 709–714
- Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nano-electrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670
- Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., *et al.* (2007) Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694
- Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P., and Heintz, N. (2014) Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* **9**, 1282–1291
- Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., and Birol, I. (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123
- Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467